Keen

911
R3L
ADM
FPK

UNITED STATES DEPARTMENT OF AGRICULTURE
COMMITTEE ON EXPERIMENTAL DESIGN
Washington 25 D. C.

THE VALUE AND USEFULNESS OF STATISTICS IN RESEARCH

by

Prof. Gertrude M. Cox, Director
Institute of Statistics
University of North Carolina
Raleigh, North Carolina

A Lecture Presented in the Department of Agriculture Auditorium
January 11, 1951

Previous Lectures Sponsored by the

Committee on Experimental Design

Dec. 14, 1949 - How Statistics Improve Physical, Chemical and Engineering Measurements. William J. Youden, Assistant Chief, Statistical Engineering Laboratory, National Bureau of Standards.

Feb. 15, 1950 - Sampling Methods in Marketing Research. Earl E. Houseman, Statistical Consultant, Bureau of Agricultural Economics.

Sept. 16, 1950 - Statistics and Research on Pasture and Grazing. Henry L. Lucas, Animal Science Statistician, North Carolina State College.

(Copies of previous lectures are available.)

PREFACE

The previous lectures in this series have been concerned with the application of the statistical method to specific fields of research. In contrast, the current lecture describes the general application of statistical principles to the several stages in the planning and execution of a research project.

Professor Cox has drawn upon her wide experience as a statistical consultant, teacher and research administrator for the principles and illustrative examples used in the lecture. She has given major emphasis to the need and value of careful planning in research. A detailed consideration of her discussion, particularly concerning the establishing of the objectives of the experiment, the role of statistics in choosing treatments, determining the size of the experiment, examining and improving techniques and in choosing the proper experimental design will profit everyone engaged in research or concerned with the administration of research projects.

This is the fourth of a series of lectures sponsored by the Departmental Committee on Experimental Design and the Agricultural Research Administration.

P. V. Cardon

Administrator
Agricultural Research Administration

THE VALUE AND USEFULNESS OF STATISTICS IN RESEARCH

By Gertrude M. Cox

I. Introduction

We shall be concerned with the usefulness of established statistical method in advancing your sciences. Some branches of science have a long record of using statistical methods while others have been slow to adopt this tool. Plant breeding is an example of a field in which biologists have been eager to take advantage of new biometric techniques, while clinical experiments with human subjects represent an area in which the penetration of statistical ideas is sporadic and incomplete.

With regard to the extent to which scientists are receptive to and utilize statistical methods, the situation is becoming more favorable. Of late years, research workers have turned increasingly to statisticians for help both in planning their experiments or surveys and in drawing conclusions from the results. This is convincing evidence that statistics has something to contribute to researc programs. The increasing acceptance of statistics as the technology of the scientific method is reflected in the rapid development of statistical laboratories and departments of statistics in our research centers and universities. In addition, one might cite the growth of a large central body of statistical theory which furnishes efficient methods of estimation and of testing hypotheses. These results account for the current vigorous output of new statistics books.

Professor Snedecor of Iowa State College thinks that people would be better adjusted to their environment if they had some knowledge of statistics. Let me quote from his new book (5), Everyday Statistics. "The common element in statistics and daily living is the element of uncertainty. Every decision in life is based on partial information so that there is always the element of uncertainty about the outcome. So it is in statistics. In studying statistics, then, you are increasing your skill in one of the basic arts of living, the art of evaluating the uncertainty of your judgments."

II. The Science of Statistics

The results of experiments are usually recorded in number symbols. The reduction of these data to summary forms with estimates of such effects as averages, variances and relationships along with tests of significance and inferences about the parameters gives rise to statistical methods.

An ideal situation would be for every experimenter to know his science and the science of statistics. However, mastery of two sciences is not possible for most of us. We shall have to advance by cooperating. In the last few years, it has been demonstrated that cooperation between statisticians and other scientists leads to better research results. It now seems that in order to secure genuine appreciation of statistics we must produce statistically minded scientists as well as well-trained statisticians.

Some statisticians feel that the lack of mathematical training is the most serious handicap of the scientist. I am inclined to be more concerned about his concepts of the scientific method. Regardless of the field of research, the scientific method requires that certain steps be taken in planning experiments. These should include (1) a statement of the objective, (2) a description of the experiment and (3) an outline of the statistical analysis of the results. The experimenter has a question which he wants answered. When the problem has been stated, then comes the outline of procedure which includes selecting the questionaire or treatment combinations; deciding on the accuracy of measurements; selecting the size, shape, number and kind of experimental units; and specifying the experimental design. Before the experiment is started, an outline of the proposed tabulations and statistical analyses should be made.

From well planned and executed experiments, the sample of observations will provide an unbiased estimate of the effects, with measures of the uncertainty of these estimates. These observations should be secured in the shortest possible time and at a minimum of cost in labor and material. I want to emphasize that if the experiment is not large enough, the answer to your question will be delayed until the experiment is repeated. If it is larger than necessary, other work that might have been conducted using the same facilities is put off until later.

Along with the three essential steps taken in planning experiments, we shall consider some of the methods used for increasing their precision. They are (1) increasing size of the experiment, (2) selecting experimental material, (3) refining techniques, (4) taking supplementary measurements and (5) choosing an efficient experimental plan. To increase the efficiency of the experiment any one or a combination of these methods may be needed to attain the desired standard of precision

III. The Role of Statistics in Experimentation

In earlier years, requests for statistical assistance were nearly always concerned with the interpretation of the results. As I have stated previously, now we are called upon to help in the planning stages. Participation in the initial stages of experiments in different areas of research leads me to a strong conviction that too little time and effort is put into the planning of research.

The statistician who expects that his contribution to the planning will involve statistical theory, finds repeatedly that he makes his most valuable contribution simply by getting the investigator to explain why he is doing the experiment, to justify the experimental treatments and to defend his claim that the experiment, when completed, will enable its objectives to be realized.

Taking into consideration the essential steps in the planning of experiments and the various methods for increasing their sensitivity, I have considered the combination of these points as an outline to illustrate how statistics can help the scientist in his research. Since the inferences that can be made depend upon the way in which the experiment is conducted, the statistician must have a detailed description of the experiment, this to include a statement of the objective; a list of the treatments selected; a description of the size of the experiment, the material, techniques and measurements to be used; an explanation of the design; and an outline of the proposed analyses.

(1) Objective. The first step in the planning of research is to state clearly defined objectives. The statement may be in the form of the question to be answered, the hypotheses to be tested, the specifications to be met, or the effects to be estimated. Because the statistician cannot evaluate the results unless he knows the objective, you will need to make your statements lucid and specific.

Let me illustrate, with an example provided by Prof. Snedecor, the importance of sound planning by means of some research results which were brought to a statistician with a question about testing the mean differences. Six protein supplements were fed to young chicks to estimate their relative effects on gain in weight. All the chicks receiving Treatment A were kept in Pen A, and similarly chicks receiving each of the other five protein supplements were kept in separate enclosures. The sexes were mixed in unknown ratios and the supplements were used in equal weights irrespective of protein content. No record was kept of individual food consumption.

Based on the statistical analysis, if the lot having Treatment A gained significantly more than the lot having Treatment B the answer might be (1) the concentration of protein in A was greater than that in B, or (2) there was a larger proportion of males in Lot A than in Lot B, or (3) the environmental conditions were more favorable in Pen A, or (4) supplement A was more appetizing than B. Such ambiguous answers mean little or no information from an experiment. This man came to the statistician to ask about the merits of various tests of significance, and he was seemingly unaware that the differences he was testing could not be identified and therefore were meaningless.

The statement of the purpose of the experiment should include an account of the area over which generalizations are to be made. Usually you will have to take a sample. The size of the sample needed depends upon the accuracy of results desired, the variability of the material being used, the magnitude of the differences to be measured, and the amount of time and money available for the study. This will be considered in more detail when we discuss the size of the experiment. However, it is well to repeatedly call your attention to the fact that the results of any piece of research, whether survey or experiment, are no better than the sample used; and as a rule a sample is interesting only if it furnishes information about the population to which the conclusions are to be applied.

In biological research, there is often a tendency to ignore the sampling problem. Suppose that the amount of some vitamin in the leaves of a nutritious plant is to be determined. The biologist may select, in a convenient way, just enough leaf material to enable the determination to be made without considering whether the material selected is a representative sample of the leaves in the field. (a) In social research sometimes a township or a town is selected for an extensive survey. The results are interpreted as if they applied to the state without having first determined whether this community was representative of the state.

(2) Treatments. After the objective of the experiment has been determined, you turn to the selection of the experimental treatments. I am using the term "treatments" in a very general sense to denote the different procedures whose effects are to be measured and compared.

In the selection of the treatments it is important to understand how they will help in reaching the objective of the experiment. Is your objective to "spot the winner" among several selections or do you want to estimate the effect of varying concentrations of a vitamin in a basic diet?

Suppose that the response to increasing amounts of the ingredient is known to be linear and that the purpose of the experiment is to determine the slope of this line. Statistics has helped to determine that for this condition the most efficient experiment contains only two levels of the ingredient. These amounts should be placed at the ends of the range within which the response is linear and the experimentation is feasible.

In certain cases the selection of the treatments has a substantial effect on the precision of the experiment. Striking gains in precision may be achieved by testing different factors in the same experiment, instead of conducting a separate experiment for each factor. This is done by means of factorial experiments in which the effects of a number of different factors are investigated simultaneously. The treatments consist of all combinations that can be formed from the different factors. These factorial experiments are often quite useful for exploratory work to see if factors have any effect, to find interactions among factors, and to use when recommendations must apply to a wide variety of conditions.

Considerable discussion may arise between the research worker and the statistician regarding the need for controls. Interesting examples can be given where the inclusion of a control has saved the experiment. As in the headache research by Jellinek (2), the control enabled isolation of those whose headaches were not relieved by an inert mixture. As he states, "discrimination among remedies for pain can be made only by subjects who have pain on which the analgesic action can be tested."

(3) Size. One of the methods for increasing the sensitivity of the experiment is to increase its size. This increase can come from increasing the number of replications by using larger experimental units. Whatever the source of the variation, replication of the experiment steadily decreases the error associated with the difference between the average effects of two treatments, provided the precautions of randomization have been taken.

How can statistics assist in determining the efficient size for the experiments?

You may need to know the number of replications necessary in order that a true difference of a given size is fairly certain (say four times out of five) to be detected as significant at the 5 percent level. Table 1 gives the number of replications required for a given probability of obtaining significant results.

From the table it is seen that if the general technique provides a standard deviation per plot of 10 percent, thirteen replications are required for a 4 out of 5 chance of detecting as significant, at the 5 percent level, a difference of 10 percent between a standard and an unknown treatment. This table should make you think if you are dealing with coefficients of variations in excess of 12 percent.

Also, it is valuable to have some idea of the sensitivity to be expected from different numbers of replications. In fact, we cannot have a high probability of detecting a 5 percent difference between the control and the test means with any reasonable number of replications unless the standard error per unit is 4 percent or under.

Table 1.- Number of replications required for a given probability of obtaining significant results.*

True difference ( ) of mean	True standard error per unit ( ) as percent of the mean: 2	4	6	8	10	12	20
5	3	9	19	33	50		
	7	22	47	--			
10	2	3	6	9	13	19	50
	3	7	13	22	33	47	--
20	2	2	2	3	4	6	13
	2	3	4	7	9	13	33
30	2	2	2	2	3	3	7
	2	2	3	4	5	7	15

Upper figure: Test of significance at 5 percent level, probability 80 percent.
Lower figure: Test of significance at 1 percent level, probability 95 percent.
*Abstracted from Table 2.1 in reference (1).

(4) Material. Two main sources of experimental variation may be distinguished. The first is lack of uniformity in the physical conduct of the experiment, or, failure to standardize the experimental technique. How statistics can help with this problem will be discussed in the next section.

The second source of variability is inherent in the experimental material to which the treatments are applied. I use the term experimental unit to denote the group of material to which a treatment is applied in a single trial of the experiment. The unit may be a plot of land, a patient in a hospital, a group of pigs in a pen, a tree, a group of plants in the pot, or one-hundred seeds in a box. It is the characteristic of such units that they produce different results even when subjected to the same treatments.

The choice of the experimental unit is important. Using uniformity data, statistics has helped in numerous fields of research to determine the best characteristics for the experimental unit. The criterion is to obtain the maximum accuracy for a given amount of money.

Statistics has helped determine the size of paddock and the number of animals per paddock for pasture research (3). "You are all aware that a very common situation in grazing work has been to use large paddocks (ones which will carry as high as 10 to 20 animals) and to use but one paddock per treatment....in recent years, paddocks which will carry only one or two animals have been employed, especially in work with dairy cattle....rather strong conflicting opinions exist as to this matter." Dr. Lucas says, "It appears that the optimum number of animals per paddock is about 7 for nutritive value studies, about 3 for yield studies with milking dairy animals, and about 2 for yield studies with other animals."

Selecting the material and evaluating its inherent variability should receive recognition as a vital part of the experiment. It is useless to spend a great deal of effort getting the chemical determinations to check within 2 percent when there is a 20 percent variation from one sample to another of the source material. At the other extreme, uniform material is sometimes selected in such a way that the responses obtained would not apply to the regular unselected material. For example, an experimental farm may be selected for its uniform soil yet the research results are to be applied to variable soil conditions. This may lead to wrong recommendations if there are strong soil x treatment interactions present.

"Selectivity of material is difficult to avoid where experiments are conducted in the field of economics and sociology. In testing some method of farm management, for instance, the experiment may require the active participation of a group of farmers. The success of such experiments depends greatly on the tact and resourcefulness of the investigator in persuading farmers to cooperate so that the participants are a representative sample of the population about which generalizations are to be drawn." /Section 2.34 reference (1)/.

(5) Technique. The other main source of experimental variation is the failure to standardize the experimental technique. When your estimated standard errors are consistently higher than those of other workers using similar material, you need to seek for the reason.

First, it might be well to list the principal objectives of a good technique. They are (1) to secure uniformity in the application of the treatments, (2) to exercise sufficient control over external influences so that every treatment produces its effect under comparable and desired conditions, (3) to devise suitable unbiased measures of the effects of the treatments and (4) to prevent gross errors.

It is in connection with improving techniques that the consulting statisticians have made one of their largest contributions to research programs.

A study was made of variability associated with the various steps in sampling cotton plots and making certain fiber measurements. The marked differences from the top to the bottom of the plant indicated the desirability of stratifying the plot samples by position on the plant.

During the testing of the effect of insecticides on the flavor of peaches, the amount of vanilla in ice cream, and other taste testing problems, designs were altered to provide ancillary technique information on the number of samples that could be evaluated by a single observer at one test period, on the effect of order of testing, and on the value of various scoring systems.

For series of experiments using radioactive materials as a tracer in tobacco and corn, a ceiling was imposed by the total number (2000) of determinations that could be handled with existing laboratory facilities. Analysis of previous data showed that the laboratory technique was much less variable than the field plots. Therefore, the experiment was designed so that the material from the six field replications were pooled by pairs giving half as many replications for the chemical determinations as for yield. If the chemical measurement is expensive, and the subsampling technique is good this compositing procedure can be very useful.

All measurements, no matter how accurate, are only approximations to some parameter. Measurement, then, is the statistical process of making inference from samples. The development of a satisfactory method of measurement may require years of research, as it has for the estimation of certain vitamins.

Investigators should be urged to consider methods of refining their techniques. It might be advisable to reconsider the purpose for which the sample is being taken and to check the intensity of sampling, the manner of compositing and sub-sampling, and the efficiency of the measurements. Also, it is worth while considering from time to time whether simplifications can be brought into the technique without undue loss of accuracy. In numerous experiments, by shifting emphasis to the major sources of variation and by reducing the amount of work done upon the less variable procedures, the total work has been reduced up to half while keeping the same efficiency.

(6) Supplementary Measurements. There is another means by which precision may, in appropriate cases, be increased by the elimination of causes of variation which cannot be experimentally controlled. In the course of an experiment it may be possible to measure related variables which predict, at least to some extent, the performances of the experimental units. Analysis of covariance enables us to estimate from the data the extent to which the results were influenced by variation in these supplementary variables.

In most dietary studies, the initial weights or total energy intakes of the children probably will affect the increases in weight during the experiment, independent of treatments. Adjustment of the observed increases to a common initial weight often greatly reduces the experimental variation. The adjusted responses represent approximately the final weights that would have been obtained if all the children had had the same initial weight.

It may be possible to take supplementary measurements which will remove variations arising from factors, as rainfall and temperature, which are impractical or impossible to control by the experimental plan.

Taking supplementary measurements has frequently meant a substantial increase in precision, increases from 20 to 70 percent in information being quite commonly realized. Of course, this depends upon how successful one has been in selecting and measuring variables which are related to the factors being investigated.

(7) Design. Finally, we attempt to minimize the experimental error by choosing an efficient design. The design represents a set of rules for allocation of the treatments to the experimental units. Each restriction has a definite purpose. The idea is to control sources of variation (as previously determined) by grouping the units on the basis of these factors. Statistics helps determine the relative importance of these sources of variation.

There are many types of designs available for your use. The simplest is the completely randomized design in which treatments are allotted to the experimental units entirely by chance. This design is frequently desirable for laboratory research, especially in physics, chemistry, bacteriology or experimental cookery, where mixing may provide a quantity of homogeneous material that can be tested under reasonably uniform conditions.

The randomized complete block design has the experimental units arranged in groups, each of which contains enough units for one set of treatments. Several such groups (replications) are needed to give an estimate of the effects being studied. This is the most commonly used design.

The latin square design carries the idea of restriction a stage further by grouping the treatments into replications in two different ways. This means that the latin square design provides more opportunity than randomized blocks for the reduction of error by skillful planning.

The incomplete block designs are adapted for experiments where a large number of treatments are to be investigated and the number of experimental units that can be considered to be homogeneous enough to fit into groups, is small.

An example will be given to illustrate how designs are selected to fit into experimental situations. In tests of mosquito repellents (6) which involve exposure of treated arms to mosquitos, the block consists of the two arms of a subject at a given time. To test 5 new repellent substances and a standard, the design given in Table 2 provides for 5 subjects, A, B, C, D, and E, each to submit his two arms to treatments on three different dates.

Table 2.- Incomplete block design

	Individuals				
	A	B	C	D	E
Date (1)	1 2	5 2	3 5	5 1	4 5
(2)	6 5	3 1	6 2	4 2	2 3
(3)	3 4	4 6	4 1	3 6	6 1

Note that every pair of treatments occurs once on some individual on one date. Repellents 1 and 2 are assigned, at random, to the right and left arms of individual A during his first test date. On the second date, his arms are submitted to repellents 6 and 5. While on the third date, his arms are treated with repellent 3 and 4. After three test dates each individual has submitted his arms to six repellents. The gain in sensitivity of incomplete block plans, as the one just illustrated, compared with randomized complete block designs, ranged from 0 to 100 percent in the mosquito repellent tests. The paired control insures utilization of gain if it is there. On an average, 5 individuals provided as much information as would have been secured from 7 or 8 individuals if the treatments had been assigned to their arms completely at random.

These incomplete block designs have been useful in biological assay, nutrition, freezing, storage and greenhouse research. In plant breeding, for hundreds of experiments, over a period of fifteen years, using 25 or more selections, the average gain in precision of these incomplete block over randomized block designs

has been about 50 percent. This means a substantial saving in time and experimental facilities when, on an average, with better designs, 4 replications are giving as good information as was secured earlier from 6 replications.

There are many types of designs available for our use. However, the limitations of the experimental material, the laboratory facilities, or the time and convenience of the investigator often dictate most of the specifications of the design. The design is the logical plan, and a well trained investigator can follow good experimental designs.

We are talking about adequate experimental designs which involve not only a satisfactory plan for conducting the trial, but include also an appropriate method for evaluating the results.

(8) Analysis. The analysis of data calls for clear thinking and for careful selection of the statistical tools to be used. It might be well to emphasize again that the statistical analysis cannot increase the validity of the data. The accuracy of mathematical inferences from data must be limited by the precision and adequacy of the observations. If you have conducted your experiment so that treatment differences are confounded with extraneous effects, statistics cannot give you reliable estimates of treatment effects.

In experiment stations and other research centers, there should be some way to get the investigators to extract all, or the major portion, of the information from their data before burying them in the files. Why should researchers be allowed to take piles of records, report a few means, file these records, and then rush out and collect more data.

An efficient, flexible set of statistical methods is available, but the use of these methods requires common sense plus an understanding of the basic assumptions involved in their use.

In analyzing results of experiments one often expects to determine relationships which are more complex than averages. Recently, statistical methods were used for the evaluation of factors affecting the incidence of syphilis. A mathematical model was constructed to enable the estimation of these factors by multiple regression. Another study which is receiving considerable thought is the relation of climate, soil type, fertilizer practice, time of harvest, methods of preparing and preserving vegetables to their nutritive properties.

Research in the quantitative aspects of genetics is making rapid strides with the assistance of statistics. In fact, statistics is an essential tool in the building of a foundation for predicting breeding and selection procedures which will yield the most rapid progress in the genetic improvement of animals and plants.

There are many statistical methods which are available for use by biologists. The newer developments in the field consist mainly of special devices such as fractional replications and sequential sampling. Sequential methods were applied to check cyst infestation rate of whitefish (4). Whitefish sometime contain cysts of the tapeworm. The presence of these cysts in the edible parts of the fish is obnoxious but not harmful to man. The inspection test used to locate the cysts is destructive, making it practical to do a minimum amount of sampling. Sequential methods meet these requirements.

In view of the growing complexity of statistical methods a trend which is unlikely to be reversed, I am sure that you will agree that more attention might profitably be given to simpler statistical methods whose efficiency is satisfactory if not the maximum attainable.

(9) Interpretation. The statistician's job is to help the experimenter plan his experiment so that the results will efficiently answer definite questions, not so they can be analyzed statistically. Then after the experiment has been conducted the results should be presented in concise and organized form so that any reasonably intelligent person can draw accurate conclusions from the results.

Interpretation of the results is the process of "making sense of figures."

IV. Conclusions

Research requires scientific knowledge, technical skill and understanding cooperation. Even then there is a risk of research wandering astray unless it is constantly being controlled by good sense.

In closing, I wish to quote from the talk Dr. Lucas presented September 6, 1950 in this "Committee on Experimental Design" series (3). "It is evident...that a proper use of statistics will enhance the efficiency of research. Statistics seems to do more than that. It cannot help but affect the attitude of the research man. It tends to make him more objective and quantitative in his work and thinking, especially on the matter of error. It tends to make him more quantitatively aware of the imperfections of experimentally determined knowledge. Thus statistics contributes to the philosophy of research."

He also spoke of the role of the statistician in research. "In addition to his role as one who knows some statistics, he is, in the biological sciences at least, aiding in the quantitative expression of fundamental knowledge about nature. This comes about through his knowledge of mathematics and his inclination to do analytical thinking. That is, in many situations, the statistician is playing the role of theorist in the applied fields."

References

1. Cochran, W. G. and Cox, G. M. Experimental Designs. John Wiley and Sons, Inc. 1950.

2. Jellinek, C. M. Clinical Tests on Comparative Effectiveness of Analgesic Drugs. Biom. Bull. 2, 87-91, 1946.

3. Lucas, Henry L. Statistics and Research on Pasture and Grazing. Lecture Series I, No. 3, U.S.D.A. Committee on Experimental Design.

4. Oakland, G. B. An Application of Sequential Analysis to Whitefish Sampling. Biometrics 6, 59-67, 1950.

5. Snedecor, G. W. Everyday Statistics. W. M. C. Brown Co., Dubuque, Iowa. 1950.

6. Wadley, F. M. Incomplete Block Design Adapted to Paired Tests of Mosquito Repellents. Biom. Bull. 2, 30-31, 1946.